Synopsis on

DESIGN OF MACHINE LEARNING ALGORITHM FOR SOLVING SOME PROBLEMS OF ENVIRONMENTAL PHYSICS

Submitted By

Sri Uday Kumar Sinha

for registration in

DOCTOR OF PHILOSOPHY
in
FACULTY OF SCIENCE

Under the supervisions of

Supervisor
Dr. Krishnanath Bandyopadhyay
Associate Professor
University Dept. of Physics
BBMK University, Dhanbad

Co-supervisor
Dr. Subhash Chandra Dutta
Head
Department of CSE & IT
BIT Sindri, Dhanbad

BINOD BIHARI MAHTO KOYALANCHAL UNIVERSITY, DHANBAD, JHARKHAND

1. Introduction

With the wide usage of Internet and smart devices/sensors by government and allied bodies for monitoring the environment, large amount of data on weather forecasting, air qualities, climate changes, natural calamities etc. are available in public domain that can be analyzed and hidden information or pattern can be revealed to predict the future events more accurately. Such large amount of data exhibit both a problem and an opportunity for scientists and environmentalists. The problem is that, it is very much difficult to analyze and process such large amount of data by human beings. The opportunity is that, this type of data is ideal for computers to process, as it is stored digitally in a well-formatted way, and it is possible to process the data much faster than that can be done by human beings [1].

Learning means ability and also procedure to acquire new knowledge or modify existing knowledge, data, behavior, skills and information. When these tasks are performed by machines without being explicitly programmed is known as machine learning, a core subarea of artificial intelligence [2]. Machine learning is something to train the computers to capable of using knowledge and behaving more intelligently than human beings do, on the basis of pattern matching, recognizing features of objects, classifying the characteristics of an object, and giving predictions by ensuring the past behaviors etc. [3].

Machine learning approach plays an important role in improving economic conditions and pattern of everyday life of the society as well as that of the country also. With the growing demand for more automation of many scientific and engineering tasks, it is necessary to make machine capable of doing tasks without being explicitly programmed using machine learning approach. The ability to apply mathematical calculations rapidly and automatically on big data to produce result accurately gives machine learning algorithm to a new level [4]. The atmospheric science and environmental physics today depend on many things such as satellite data processing for weather forecasting and climate change, analyses of air and water quality, noise pollution, radiation hazards and global warming etc. [5] where use of machine learning approaches can significantly reduce the manual dependence of the scientists / researchers.

As the machine learning algorithm works with huge data set, traditionally, if it allow to be executed in single processor environment the algorithmic bottlenecks can lead to substantial delays in model processing, from training, to classification, to distance and error calculations, and beyond. Therefore many parallel algorithms and architectures have been designed to address these issues such as parallel forecasting [6], parallel prefix algorithms [7], data concentration [8], GP-GPU [9], CUDA parallel programming framework [10] etc. which applied in some environmental physics problem like air pollutions [11], Nuclear, Energy, and Environmental Applications [12], hurricane modeling [13] etc.

Environmental conditions are always subject to changes due to periodical (sometimes a periodical) changes in seasons and change in geographical positions. These changes need to be predicted accurately to make the lives easier for people by giving proper forecasting and warning signals. The wrong information could be disastrous to civilians and military especially for aircrafts and military operations while the good forecasting may assist them and the government to take proper steps. A good weather forecasting scheme may also help farmers to plan crop irrigation. The business and activities like boating, fishing, skimming, scuba diving etc. are also directly dependent on the weather conditions [14-19]. Such type of activities can be possibly monitored using machine learning approaches.

2. Literature Review

Artificial Intelligence is a kind of systems using a form of knowledge representation to maintain data with some objectives and that data may or may not be dynamically altered and it is intended to achieve the solutions in unknown environment. Machine learning is a subset of Artificial Intelligence which uses some statistical methods to enable the machine to improve its functions with human beings, whereas Deep Learning is a subset of Machine Learning which basically uses multi-layer perceptions like neural networks of human minds.

In recent years, several models based on machine learning techniques have been developed to predict environmental issues [4] based on the data available from different organizations. Among the commonly used machine learning techniques, Artificial Neural Networks (ANN)[5], Support Vector Machines (SVM)[14], Decision Trees (DT) [15] etc. are the most popular among them.

Machine learning tools can be used for predicting water quality [16], air pollution [17-20], noise pollution etc. with maximum possible accuracy based on their previous applications. Similarly catastrophic environmental issues like thunderstorms [13], flood [22,23], rainfall [24-28], global warming [29-31], soil moisture [31-33], sand storms [34], radiation hazard [35-36] can also be addressed successfully by machine learning taking into considerations the phenomena of solar irradiation, radiation oncology and global temperature.

3. Proposed Possible Problems

Keeping the above points in mind, the plan of the proposed research work can be summarized as given below .We would like to carry out research works in the following domain(s):

1. An extensive analytical as well as comparative study of the existing applications and algorithms have to be done considering the problems of environmental physics, particularly climate change, weather forecasting, rainfall, radiation hazards etc. For this we have to develop some algorithms/models for analyses and accordingly suggestions / predictions of these environmental issues would be made by using machine learning approach.

2. The work can further be extended to discuss other environmental problems such as environmental pollution (air pollution, water pollution, noise pollution etc.) and disease control by using machine learning algorithm.
3. A comparative study of machine learning methods and feature engineering techniques in terms of improved predictions for the above problems of environmental physics might also be done.

4. Objective(s)

Basic objective of the work is to be familiar with various factors influencing the problems of environmental physics that pervade the environment in our everyday lives. To get relief from or get rid of these issues we have to design and develop some models using machine learning approach to monitor environmental issues or problems as mentioned above. The machine learning algorithm would undoubtedly help to determine future trends of lifesaving decisions such as

disaster preparedness and recovery from natural calamities and other possible factors influencing the environment.

5. Methodology

In order to carry out the proposed research work we have to start with the various tools and models of machine learning that can be mapped over the data available on the problems of environmental physics as mentioned above. Following methodology may be followed from time and again:

- To develop some models and algorithms using machine learning approach and to train the system to generate some results after studying above mentioned problems. Further, the developed models would be analyzed to calculate the possible errors in the results obtained and subsequent steps (if necessary) may be applied to make corrections in models so that errors should be least.
- To collect data from environment monitoring agencies/concerned departments. Data related to rain fall, air pollution, water pollution, radiation hazards, disease control etc. of Dhanbad and nearby districts of Jharkhand would be given topmost priority. However, the above mentioned problems might be studied on the national basis, if possible.
- In order to examine the mapping of data in machine learning environment, consultations with nearby scientific bodies working in the area of environmental sciences such as IIT(ISM), CIMFR or other research institutions may be arranged.

6. Tools / Mechanism

Machine learning provides capability to the systems to learn and improve from experience without being programmed by a user. With the introduction of machine learning concept, there is a great ease for data analysis and prediction. Machine learning does not require understanding of the physical processes that control the atmosphere / environment but uses the past data to predict future results. Therefore, this process may be used for weather forecasting, global warming, environmental pollutions and natural calamities. Computer-aided predictions on weather, pollutions and natural calamities are becoming rapidly developing field of pattern

recognition and machine learning applications with strong potential to be applied in the near future.

The available tools such as Python, MATLAB, Deep Learning etc. may be used to perform the proposed research work.

7. Significance/Importance of proposed work

The proposed work is closely related to environment which is directly or indirectly affecting everyday life. Thus, the importance of proposed work may be summarized as follows.

- Some suggestions or predictions regarding natural calamities like climate change, heavy rainfall / less rainfall, cyclone, drought, rise in sea level due to global warming, noise pollutions, radiation hazards, disease control etc. could be made using machine learning approach,
- And accordingly, the government as well as different institutions or bodies can take necessary measures either to avoid these problems or to reduce the impact of such calamities.
- Finally, machine learning approach for solving the real life problems of human beings as well as habitants of the society might give relief to them.

8. Limitations

Any system cannot be perfectly designed and thus, there is always a chance for improvement in its functions. Therefore, the models or algorithms to be developed during the work would always be tested for its working in adverse environment. Limitations of the developed models should be analyzed after calculating the errors by comparing the past records or data available so far and, hence, the models can be refined /modified to minimize the errors in case of future predictions.

Again, collection of accurate data is also a challenging task. As the proposed work is directly dependent on the quality of data hence, best efforts should have to be given to refine the data in best possible way, but there always remains a chance of error in data processing. However, emphasis on choice of parameters as well as data collection should be given to make the error least.

9. Tentative Chapters

The tentative chapters of the proposed research work may be classified as follows:

Chapter 1: Introduction

(i) Preliminary review of various works carried out during the past few years using machine learning approach.

(ii) Discussions of models and algorithms that have been used so far and also in the present work to get insight for solving the problems as proposed.

(iii) About the source(s) of data.

(iv) Plan of the thesis

Chapter 2: Methodology Used

Methodology adopted for analyzing the data collected for the present purpose (mainly programming language(s) to be used for getting most effective solutions for the proposed problems).

Chapter 3: Problem – I

Comprehensive study of the problems on climate change, weather forecasting , rain fall etc. will be presented in this chapter (in the context of Dhanbad and some other nearby districts of Jharkhand state; if possible, some other states may be considered).

Chapter 4: Problem – II

Exhaustive study and thorough analyses on research works to be carried out for air pollution, water quality, noise pollution and radiation hazards as far as possible will be elaborated in this chapter considering selected places/bodies of Dhanbad district.

Chapter 5: Result and Discussions

Results of research works as mentioned in Chapters 3 and 4 would be summarized and in-depth discussions will be made available in this chapter.

Chapter 6: Conclusions

This chapter will contain the conclusions of the research works proposed in the thesis.

References:

All relevant references and available data link will be placed in this section as per guidelines of the university.

References

1. S. Ray, "A Quick Review of Machine Learning Algorithms," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, 2019, pp. 35-39.
2. P. Pandey and R. Prabhakar, "An analysis of machine learning techniques (J48 &AdaBoost)-for classification," 2016 1st India International Conference on Information Processing (IICIP), Delhi, 2016, pp. 1-6.
3. S. E. Haupt, J. Cowie, S. Linden, T. McCandless, B. Kosovic and S. Alessandrini, "Machine Learning for Applied Weather Prediction," 2018 IEEE 14th International Conference on e-Science (e-Science), Amsterdam, 2018, pp. 276-277.
4. N. Singh, S. Chaturvedi and S. Akhter, "Weather Forecasting Using Machine Learning Algorithm," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 171-174.
5. T. R. V. Anandharajan, G. A. Hariharan, K. K. Vignajeth, R. Jijendiran and Kushmita, "Weather Monitoring Using Artificial Intelligence," 2016 2nd International Conference on Computational Intelligence and Networks (CINE), Bhubaneswar, 2016, pp. 106-111.
6. S. K. Jha and P. K. Jana, "Parallel algorithm for time series based forecasting on OTIS-Mesh", International Journal of Computer Applications, vol. 26(1), page 70-75, 2010.
7. S. K. Jha and P. K. Jana, "Fast parallel prefix on multi-mesh of trees," *2010 International Conference on Computer and Communication Technology (ICCCT)*, Allahabad, Uttar Pradesh, 2010, pp. 641-646,
8. U. K. Sinha, S. K. Jha and H. Mandal, Data Concentration on Torus Embedded Hypercube Network. Advanced Computing, Networking and Informatics- Volume 2. Smart Innovation, Systems and Technologies, vol 28. Springer, Cham, 2014 pages 525-532,
9. Xiaohui Wei, Hengshan Yue, Shang Gao, Lina Li, Ruyu Zhang, Jingweijia Tan, G-SEAP: Analyzing and characterizing soft-error aware approximation in GPGPUs, Future Generation Computer Systems, Volume 109, 2020, Pages 262-274

10. Jorge González-Domínguez, Roberto R. Expósito, Verónica Bolón-Canedo, CUDA-JMI: Acceleration of feature selection on heterogeneous systems, Future Generation Computer Systems, Volume 102, 2020, Pages 426-436,
11. K. Georgiev and Z. Zlatev, Parralel Computing with Large Scale Computations in Air Pollution Modelling. NATO Science Series (Series 2: Environmental Security), vol 57. Springer, Dordrecht, 1999
12. D. Gaston, L. Guo, G. Hansen, H. Huang, et. al. Parallel Algorithms and Software for Nuclear, Energy, and Environmental Applications. Part II: Multiphysics Software. Communications in Computational Physics, 12(3), 834-865, 2012.
13. Ali Abdolali, Aron Roland, Andre van der Westhuysen, Jessica Meixner, Arun Chawla, Tyler J. Hesser, Jane M. Smith, Mathieu Dutour Sikiric, "Large-scale hurricane modeling using domain decomposition parallelization and implicit scheme implemented in WAVEWATCH III wave model", Coastal Engineering, Volume 157, April 2020, 103656
14. N. L. and M. H.S., "Atmospheric Weather Prediction Using various machine learning Techniques: A Survey," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 422-428.
15. S. S. Bhatkande and Roopa G. Hubballi "Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques," International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 5, May 2016.
16. M. Bilal, A. Gani, M. Marjani and N. Malik, "A Study on Detection and Monitoring of Water Quality and Flow," 2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), Karachi, Pakistan, 2018, pp. 1-6.
17. S. Ameer et al., "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities," in IEEE Access, vol. 7, pp. 128325-128338, 2019.
18. T. M. Amado and J. C. Dela Cruz, "Development of Machine Learning-based Predictive Models for Air Quality Monitoring and Characterization," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 0668-0672.
19. C. Srivastava, S. Singh and A. P. Singh, "Estimation of Air Pollution in Delhi Using Machine Learning Techniques," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, Uttar Pradesh, India, 2018, pp. 304-309.
20. S. Mahanta, T. Ramakrishnudu, R. R. Jha and N. Tailor, "Urban Air Quality Prediction Using Regression Analysis," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 1118-1123.
21. Y. Jiang, J. Yao and Z. Qian, "A Method of Forecasting Thunderstorms and Gale Weather Based on Multisource Convolution Neural Network," in *IEEE Access*, vol. 7, pp. 107695-107698, 2019.

22. M. Khalaf, A. J. Hussain, D. Al-Jumeily, P. Fergus and I. O. Idowu, "Advance flood detection and notification system based on sensor technology and machine learning algorithm," 2015 International Conference on Systems, Signals and Image Processing (IWSSIP), London, 2015, pp. 105-108.
23. M. Khalaf et al., "A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, 2018, pp. 1-8.
24. C. Thirumalai, K. S. Harsha, M. L. Deepak and K. C. Krishna, "Heuristic prediction of rainfall using machine learning techniques," 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, 2017, pp. 1114-1117.
25. V. Abhyankar, A. G. Singh, P. Paul, A. Mehta and S. Vidhya, "Portable Autonomous Rain Prediction Model Using Machine Learning Algorithm," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-4.
26. A. Kala and S. G. Vaidyanathan, "Prediction of Rainfall Using Artificial Neural Network," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 339-342.
27. M. P. Darji, V. K. Dabhi and H. B. Prajapati, "Rainfall forecasting using neural network: A survey," 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, 2015, pp. 706-713.
28. U. Shah, S. Garg, N. Sisodiya, N. Dube and S. Sharma, "Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan Himachal Pradesh, India, 2018, pp. 776-782.
29. C. Thirumalai, G. Saikrishna, C. S. Raju and M. Senthilkumar, "Analysis of global warming in India over maximum temperature using Pearson and machine learning," 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, 2017, pp. 496-499.
30. H. S. Kaur and S. Randhawa, "Global Land Temperature Prediction by Machine Learning Combo Approach," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, 2018, pp. 1-8.
31. K. Bandyopadhyay, “Solar Activity, Global Warming and Climate Change : A Brief Review”, Proc. Of Challenges in Weather Forecasting, eds. D. Lohar et al., ACB Publications, (ISBN: 81-87500-60-3), Kolkata, India, 2012, pp. 135 - 141 .
32. S. Prakash, A. Sharma and S. S. Sahu, "Soil Moisture Prediction Using Machine Learning," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1-6
33. S. Chatterjee, S. Kumar, J. Saha and S. Sen, "Hybrid Regression Model for Soil Moisture Quantity Prediction," 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), Kolkata, India, 2019, pp. 1-5.

34. H. A. Shaiba, N. S. Alaashoub and A. A. Alzahrani, "Applying Machine Learning Methods for Predicting Sand Storms," 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, 2018, pp. 1-5.
35. S. Sanders, C. Barrick, F. Maier and K. Rasheed, "Solar Radiation Prediction Improvement Using Weather Forecasts," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, 2017, pp. 499-504.
36. E. D. Obando, S. X. Carvajal and J. Pineda Agudelo, "Solar Radiation Prediction Using Machine Learning Techniques: A Review," in IEEE Latin America Transactions, vol. 17, no. 04, pp. 684-697, April 2019.
37. K. Bandyopadhyay, “Mobile Communication: Unpleasant Effects – A Brief Review”, Proc. Of the National Seminar on Effects of Mobile Phone Radiation on Plant, Human & Other Animals, Sponsored by Vinoba Bhave University, Hazaribag, Jharkhand and Organised by Department of Physics & Botany, Chas College, Bokaro, Jharkhand, India, pp. 38, April 23 -24, 2016.